ISKO BRASIL
International Society for Knowledge Organization
CANELA, RS 2025

Revisão narrativa da literatura sobre o uso da IA generativa na construção e manutenção de taxonomias

Generative AI in the construction and maintenance of taxonomies: a narrative review

Amanda Souza (1), Elisangela Agnette (2)

(1) Universidade Federal de Minas Gerais, Minas Gerais, amanda.jercika@gmail.com, ORCID: 0009-0004-1026-0065
(2) elisangelaaganette@gmail.com, ORCID: 0000-0003-4357-8016

DOI: 10.22477/ISKO25.101

Resumo

As taxonomias são estruturas essenciais para a organização e recuperação do conhecimento em ambientes digitais. No entanto, sua construção e atualização manuais são complexas e custosas, especialmente diante do crescimento exponencial de dados. A inteligência artificial generativa surge como uma alternativa promissora para automatizar e otimizar esse processo. **Objetivo**: Este estudo analisa como a inteligência artificial (IA) generativa pode otimizar a construção e reformulação de taxonomias em contextos organizacionais. **Metodologia**: Adota-se uma abordagem qualitativa, com revisão narrativa da literatura, análise de conteúdo e estudo dos métodos utilizados em 19 artigos selecionados entre 792 publicações recuperadas entre 2016 e 2025. **Resultados**: A pesquisa identificou quatro enfoques principais: indução de taxonomias, enriquecimento, avaliação e geração de conceitos, com destaque para o uso de LLMs e técnicas como *prompting* estruturado e aprendizado profundo. Foram observados benefícios como escalabilidade e adaptação dinâmica, mas também desafios, como ambiguidade semântica e ausência de curadoria humana. **Conclusão**: A IA generativa representa um avanço promissor na gestão da informação, mas sua eficácia depende da combinação com a expertise de profissionais da informação, garantindo qualidade, coerência e relevância das taxonomias geradas.

Palavras-chave: Taxonomias. Inteligência artificial generativa. Organização do conhecimento.

Abstract

Taxonomies are essential structures for organizing and retrieving knowledge in digital environments. However, their manual development and maintenance are complex and costly, especially given the exponential growth of data. Generative artificial intelligence emerges as a promising alternative to automate and enhance these processes. **Objective:** This study analyzes how generative artificial intelligence (AI) can optimize the construction and refinement of taxonomies in organizational contexts. **Methodology:** A qualitative approach was adopted, including a narrative literature review, content analysis, and examination of methods used in 19 selected articles from a pool of 792 publications (2016–2025). **Results:** The study identified four main approaches: taxonomy induction, enrichment, evaluation, and concept generation, highlighting the use of LLMs, structured prompting, and deep learning. While benefits such as scalability and dynamic adaptation were found, challenges remain, including semantic ambiguity and lack of human curation. **Conclusion:** Generative AI presents a promising advance in knowledge organization. However, its effective implementation requires integration with the expertise of information professionals to ensure quality, coherence, and relevance of the generated taxonomies.

Keywords: Taxonomies. Generative artificial intelligence. Knowledge organization.

Resumen

Las taxonomías son estructuras fundamentales para organizar y recuperar conocimiento en entornos digitales. No obstante, su construcción y mantenimiento manual resultan complejos y costosos, especialmente ante el crecimiento exponencial de los datos. La inteligencia artificial generativa se perfila como una alternativa eficaz para automatizar y mejorar estos procesos. **Objetivo:** Este estudio analiza cómo la inteligencia artificial (IA) generativa puede optimizar la construcción y reformulación de taxonomías en contextos organizacionales. **Metodología:** Se utilizó un enfoque cualitativo basado en revisión narrativa de la literatura, análisis de contenido y evaluación de métodos en 19 artículos seleccionados de un total de 792 publicaciones (2016–2025). **Resultados:** Se identificaron cuatro enfoques principales: inducción, enriquecimiento, evaluación y generación de conceptos taxonómicos. Se destaca el uso de LLMs, prompting estructurado y aprendizaje profundo. Entre los beneficios se incluyen la escalabilidad y adaptación dinámica; sin embargo, persisten desafíos como la ambigüedad semántica y la ausencia de validación humana. **Conclusión:** La IA generativa representa un avance significativo en la organización del conocimiento. No obstante, su aplicación eficaz requiere la colaboración con expertos en información para garantizar la calidad, coherencia y pertinencia de las taxonomías generadas.

Palabras clave: Taxonomías. Inteligencia artificial generativa. Organización del conocimiento.

1 Introdução

Na sociedade contemporânea, pautada pelo conhecimento, a gestão da informação e do conhecimento configura-se como um recurso estratégico essencial para as organizações, permitindo-lhes obter vantagens competitivas ao potencializar suas experiências internas e responder com precisão às demandas de seus clientes (Trevisan; Damian, 2018). Nesse contexto, estruturas como as taxonomias desempenham papel fundamental ao sistematizar e organizar o conhecimento por meio de categorias hierárquicas, viabilizando a classificação de documentos e conteúdos organizacionais. Para que sejam eficazes, tais taxonomias devem ser desenvolvidas de maneira flexível, pragmática e coerente, de modo a se adaptarem às necessidades específicas de cada organização (Woods, 2004).

Este estudo tem como objetivo analisar, a partir da literatura recentemente publicada em diversas áreas do conhecimento, incluindo a Ciência da Informação, como os modelos e técnicas de IA generativa podem contribuir para a construção e reformulação de taxonomias.

O desafio deste estudo reside no fato de que a construção e reformulação de taxonomias possui uma natureza complexa, principalmente em cenários com grande quantidade de dados, o processo manual pode tornar-se insustentável ou em um curto período tornar a taxonomia obsoleta (Andriani, 2017). A Inteligência Artificial (IA) generativa tem o potencial de facilitar esse processo ao identificar padrões e sugerir reformulações automáticas, viabilizar uma categorização flexível, propiciar a construção dinâmica dos relacionamentos dos conteúdos e otimizar o processo de reformulação das taxonomias. No entanto, existem desafios consideráveis, incluindo a confiabilidade das classificações geradas automaticamente, a presença de vieses nos algoritmos, qualidade dos dados, falta de contexto, ambiguidade semântica e necessidade da interação humana.

Considerando o cenário apresentado, este estudo possui a seguinte questão de pesquisa: de que maneira a inteligência artificial generativa pode ser empregada na otimização da construção e reformulação de taxonomias?

As motivações para realização deste estudo estão direcionadas às reflexões acerca da inteligência artificial generativa, no contexto da ciência da informação, com enfoque para as taxonomias.

No campo da Ciência da Informação, o uso de IA generativa representa uma ampliação das abordagens tradicionais de organização do conhecimento, ao integrar modelos computacionais avançados que podem aprimorar a eficiência e a adaptabilidade das estruturas classificatórias. Já do ponto de vista das grandes empresas, a adoção de IA generativa para a reformulação de taxonomias têm implicações diretas na competitividade das organizações, especialmente aquelas que operam com grandes volumes de dados, como empresas de tecnologia, consultorias entre outras. Entende-se que esse novo contexto abre um campo fértil para oportunidades de pesquisa e inovação, tanto do ponto de vista teórico quanto prático para a área da Ciência da Informação

2 Referencial teórico

Este capítulo apresenta uma contextualização concisa da temática da pesquisa, fundamentando-se em referencial teórico para embasar a análise dos resultados à luz do objetivo estabelecido.

2.1 Taxonomias

O conceito de taxonomia tem origem na classificação biológica proposta por Carl Linnaeus no século XVIII, que estabeleceu um sistema hierárquico para categorizar organismos. Essa abordagem influenciou várias áreas, incluindo a ciência da informação e a gestão do conhecimento (Mayr, 1969). Na educação, Benjamin Bloom (1956) desenvolveu a Taxonomia dos Objetivos Educacionais, que foi atualizada por Anderson e Krathwohl (2001) para refletir uma compreensão mais dinâmica da aprendizagem.

Na ciência da informação, Ranganathan (1933) introduziu a classificação facetada, que trouxe maior flexibilidade ao organizar o conhecimento. Aganette (2010) propôs uma abordagem adaptativa às taxonomias, destacando a importância de considerar a interconectividade e a dinâmica das categorias no contexto digital emergente. Em 2023, Aganette e Maculan exploraram as lições derivadas de uma taxonomia fictícia de Borges, abordando os desafios na construção de sistemas de classificação e a importância da clareza hierárquica.

Taxonomia pode ser definida como um sistema que organiza elementos em categorias com base em características comuns (Nickerson *et al.*, 2009). Na ciência da informação, ela facilita a indexação e recuperação de dados (Rowley, 1992), enquanto na educação organiza habilidades em níveis progressivos (Bloom, 1956). Na gestão do conhecimento, auxilia na classificação dos ativos intelectuais de uma organização (Bailey, 2012).

Autoria contemporânea, como Smith e Johnson (2023), propõe que as taxonomias digitais sejam dinâmicas, atualizadas constantemente para refletir a evolução de conceitos e práticas. Tecnologias como IA e aprendizado profundo são usadas para aprimorar a categorização de dados (González; Pereira, 2022).

A taxonomia se caracteriza por elementos como hierarquia, facetas e exclusividade (Mccarthy, 1995). No ambiente digital, ela se distingue das ontologias, que permitem relações semânticas complexas, e das folksonomias, que são baseadas em classificações colaborativas (Gruber, 1993; Shirky, 2005). Pesquisas recentes, como as de Brown e Silva (2023), indicam que a taxonomia digital deve ser complementada por análise semântica automatizada e aprendizado de máquina para uma categorização mais precisa.

Portanto, a taxonomia permanece essencial para a organização do conhecimento, evoluindo conforme as necessidades dos diferentes campos do saber.

2.2 Inteligência Artificial e Inteligência Artificial Generativa

A inteligência artificial (IA) tem suas raízes em diversas disciplinas como filosofia, psicologia, matemática e linguística, que contribuíram com ideias, perspectivas e técnicas para o desenvolvimento de sistemas inteligentes. Embora os primeiros trabalhos sobre IA tenham ocorrido entre 1945 e 1955, foi na conferência de Dartmouth, em 1956, que a IA se consolidou como campo de estudo. Nesse evento, uma nova geração de pesquisadores traçou as diretrizes e perspectivas futuras, alternando entre novas abordagens criativas e o aprimoramento das mais promissoras (Ruiz *et al.*, 2003; Russell; Norvig, 2016).

Definir IA é desafiador, como apontado por Stuart Russell e Peter Norvig no livro “Inteligência Artificial”. Inicialmente, eles afirmam que a IA é um tema interessante, mas sem uma definição precisa. Em seguida, o livro apresenta várias definições, sendo algumas baseadas em processos de pensamento e comportamento humano, e outras em sistemas racionais mais abstratos e formais. Essas abordagens podem ser classificadas em quatro categorias principais: sistemas que pensam como humanos, sistemas que agem como humanos, sistemas que pensam racionalmente e sistemas que agem racionalmente. A busca por uma inteligência semelhante à humana envolve, por um lado, uma abordagem empírica relacionada à psicologia e, por outro, uma abordagem racionalista que combina matemática e engenharia, com conexões à estatística e economia (Russell; Norvig, 2016).

De forma geral, um agente inteligente deve ser capaz de representar conhecimento, lidar com incertezas, raciocinar, tomar decisões, aprender com experiências e instruções, e interagir tanto com outros agentes quanto com o ambiente (Cozman; Neri, 2021). A IA abrange diversos subcampos e é objeto de estudo de várias áreas, cada uma voltada para objetivos específicos que contribuem de forma inovadora para o avanço dessa tecnologia. Destaca-se a Ciência da Computação, que tem registrado grandes avanços, especialmente com a ascensão das Inteligências Artificiais Generativas (GAI). Essas IAs geram conteúdos como textos, imagens, músicas ou códigos com base em grandes volumes de dados e instruções humanas (prompts) (Coneglian *et al.*, 2023; Casar Corredera, 2023).

A partir do final de 2022 e ao longo de 2023, as IAs generativas, impulsionadas pelos *Large Language Models* (LLMs), como ChatGPT, Copilot, Gemini e DeepSeek, provocaram uma transformação significativa no mercado e na academia. Esses modelos redefiniram o papel da IA e expandiram suas aplicações, além de reformularem as perspectivas sobre seu impacto em diversas áreas do conhecimento. A tendência é que, nos próximos anos, o impacto da IA generativa cresça ainda mais, com soluções especializadas em áreas como saúde, educação, gestão e organização do conhecimento.

3 Metodologia

Para caracterizar e delimitar este estudo, adotaram-se critérios propostos por Gil (2008) e Prodanov e Freitas (2013), uma vez que esses parâmetros orientam a definição da estrutura e dos objetivos. A pesquisa adota uma abordagem qualitativa, de natureza exploratória, com base em uma revisão narrativa e análise de casos de uso. Essa metodologia permite compreender como modelos e técnicas de Inteligência Artificial (IA) generativa podem contribuir para a reformulação de taxonomias corporativas. O estudo será conduzido em três fases principais, descritas a seguir.

Fase 1 - Revisão Narrativa da Literatura

Na primeira fase deste estudo, foi adotada uma revisão narrativa da literatura, uma vez que o objetivo consistiu em mapear, de forma abrangente, crítica e interpretativa, as contribuições acadêmicas sobre o uso da IA generativa na construção e no refinamento de taxonomias. As etapas operacionais seguiram a seguinte estrutura: (i) definição do problema de pesquisa e da questão orientadora; (ii) seleção das bases de dados acadêmicas e repositórios científicos relevantes, como *Google Scholar, Scopus* e *Web of Science*; (iii) definição das palavras-chave específicas, incluindo "*taxonomy refinement*", "*adaptive taxonomy*", "*dynamic taxonomy generation*", "*taxonomy evolution*", "*automatic taxonomy*", "*intelligence taxonomy*", "*automatic taxonomy generation*", "*generative AI*", "*generative artificial intelligence*", "*machine learning*", "*neural networks*" e "*large language models*"; (iv) aplicação de filtros que delimitam publicações entre 2016 e 2025, restringindo-se aos idiomas inglês, português e espanhol, publicações *Open Access*, além de excluir estudos provenientes das áreas de biologia, zoologia, botânica e educação; (v) leitura exploratória e seleção de artigos significativos; (vi) análise crítica e categorização temática dos achados; e (vii) síntese dos resultados em quadro analítico (Quadro 1). A escolha pela revisão narrativa justifica-se pela necessidade de flexibilidade metodológica, essencial para abarcar diferentes perspectivas teóricas e aplicações práticas em campos ainda em rápida evolução, como o da Inteligência Artificial aplicada à Organização do Conhecimento.

Fase 2 - Consolidação dos Resultados e Análise de Conteúdo

Com os documentos coletados na fase anterior, será realizada uma análise de conteúdo para selecionar os estudos mais relevantes à temática da pesquisa. Essa análise será conduzida de forma sistemática, identificando os principais métodos e técnicas utilizados na construção e reformulação de taxonomias por meio da IA generativa. Os critérios de inclusão para os documentos analisados consideraram a pertinência ao tema e a solidez metodológica do estudo. Essa etapa permitirá uma categorização dos diferentes enfoques teóricos e metodológicos encontrados na literatura.

Fase 3 - Análise dos Métodos e Técnicas de IA Generativa Aplicadas à Reformulação de Taxonomias

Na terceira fase, será realizada uma análise descritiva dos métodos e técnicas de IA generativa identificados nos textos selecionados. Essa análise visa compreender de que maneira essas abordagens podem ser aplicadas para a construção e reformulação de taxonomias, destacando vantagens, desafios e oportunidades. Serão consideradas técnicas como redes neurais generativas, modelos de linguagem de grande escala (LLMs), aprendizado profundo e algoritmos de *clustering*. Além disso, serão examinados casos de uso e aplicações práticas que demonstrem o impacto dessas tecnologias na gestão do conhecimento e na organização da informação dentro das empresas.

Acredita-se que a combinação dessas três fases permitirá obter um panorama abrangente sobre a aplicação da IA generativa na otimização de taxonomias.

4 Resultados e discussões

Fase 1 - Revisão narrativa de literatura

A revisão realizada nas bases científicas definidas na metodologia resultou na identificação de 792 publicações, incluindo artigos de periódicos, artigos de revisão e trabalhos apresentados em eventos. A análise desse material revelou que, nos últimos dez anos, a maior concentração de estudos sobre a temática investigada ocorreu na área de Ciência da Computação. Quanto ao formato das publicações recuperadas, verificou-se que a maioria corresponde a artigos de periódicos e trabalhos apresentados em eventos, indicando uma produção científica consolidada e de alto impacto.

Durante o processo de seleção, foram aplicados filtros e critérios de inclusão e exclusão para garantir a relevância dos estudos. Publicações duplicadas foram eliminadas, assim como estudos das áreas biológicas que, apesar dos filtros aplicados nas bases de dados *Scopus* e *Web of Science*, ainda foram recuperados.

Além disso, ao analisar os títulos, palavras-chave e resumos, observou-se que muitos dos trabalhos recuperados não abordavam diretamente o uso de IA generativa na reformulação de taxonomias. Embora esses estudos mencionaram taxonomia ou inteligência artificial, aqueles que não apresentavam métodos ou técnicas essenciais para modelos generativos foram descartados. Dessa forma, 83 estudos foram selecionados, distribuídos entre as bases de dados da seguinte maneira: I) *Web of Science* (*WoS*): 4 estudos; II) *Scopus*: 5 estudos e III) *Google Scholar*: 74 estudos

Para identificar tendências de pesquisa e relações entre os temas, foi realizada uma análise de palavras-chave utilizando o software *VOSviewer*. Definiu-se um número mínimo de duas ocorrências por termo e utilizou-se a normalização por força de associação para destacar as conexões mais relevantes.

A análise do mapa gerado revelou a formação de oito clusters, evidenciando a interconectividade e a evolução das pesquisas sobre taxonomias e IA generativa. O termo de maior ocorrência e centralidade foi *Natural Language Processing* (NLP), associado a 12 termos e apresentando a maior força de conexão no mapa (20 links). Esses resultados reforçam a relevância do NLP na organização do conhecimento, sendo um dos pilares fundamentais da aplicação de IA generativa em taxonomias.

O termo "taxonomia", como esperado, também se destacou, associando-se a oito outros termos, com uma predominância de publicações em 2021. Suas conexões refletem a evolução do campo, apontando para avanços na automação e enriquecimento de taxonomias, um tema que ganhou maior relevância a partir de 2023.

A Ciência da Informação apareceu no mapa conectada a publicações de 2021, evidenciando o crescente interesse da área na automatização da construção e reformulação de taxonomias. Essa abordagem está alinhada a técnicas como *Machine Learning, Word Embedding e Large Language Models* (LLMs), que também ocuparam uma posição central na análise.

A ascensão dos LLMs a partir de 2023 pode estar diretamente relacionada à popularização da IA generativa, como por exemplo o ChatGPT (Casar Corredera, 2023). Os termos de menor ocorrência podem sugerir tendências emergentes, como: I) Automação do enriquecimento de taxonomias e II) Alinhamento de taxonomias. Ambos os conceitos são cruciais para garantir compatibilidade semântica entre diferentes estruturas taxonômicas, promovendo interoperabilidade e otimização da recuperação da informação.

Por fim, destaca-se que, embora o avanço da IA generativa seja evidente nas conexões entre termos e técnicas, nenhum dos trabalhos analisados utilizou explicitamente a expressão "IA generativa" como palavra-chave. Esse achado sugere que as pesquisas ainda se concentram na exploração e desenvolvimento de técnicas subjacentes, como NLP, *Machine Learning*, LLMs e *Word Embedding*, que fundamentam esse novo paradigma tecnológico.

Em relação à evolução da quantidade de publicações na última década (2016 a 2025), observou-se uma flutuação na produção científica sobre o tema. Os anos de 2017, 2019 e 2022 apresentaram um declínio no número de publicações, enquanto 2018, 2020 e 2021 indicaram uma tendência de crescimento. Esse comportamento sugere uma emergência do tema, impulsionada por avanços tecnológicos recentes e pelo crescente interesse no potencial da IA generativa.

Fase 2 - Consolidação dos Resultados e Análise de Conteúdo

Com base nos documentos coletados na fase anterior, foi realizada uma análise sistemática para identificar os principais métodos e técnicas empregados na aplicação de taxonomias mediadas por inteligência artificial generativa. A seleção dos artigos considerou sua relevância temática e contribuição teórico-metodológica para a compreensão das interseções entre taxonomias e IA generativa. Como resultado desse processo, foram selecionados 19 artigos.

Para auxiliar o processo de seleção, utilizou-se o modelo de linguagem ChatGPT-4 para a geração de resumos analíticos dos artigos, com o apoio do *framework* RISE (Papel - Entrada - Passos - Expectativas). Esse *framework* orienta a formulação de prompts de maneira estruturada, fornecendo contexto, detalhando etapas e estabelecendo expectativas claras quanto ao formato e à profundidade das respostas geradas. A utilização dessa abordagem visou otimizar a compreensão inicial dos textos, sem substituir a análise humana. Após a geração dos resumos, os artigos foram lidos criteriosamente, com o intuito de garantir a acurácia das informações extraídas e assegurar a adequação dos documentos aos critérios definidos para a seleção final.

A análise dos artigos revelou o uso de IA generativa com diferentes níveis de sofisticação, abrangendo desde LLMs contemporâneos até modelos clássicos, como seq2seq e sistemas estatísticos, a exemplo do Moses. Essa diversidade reflete o recorte temporal de dez anos adotado no estudo. Ainda que menos avançados, os modelos clássicos também geram linguagem natural e, por isso, são considerados formas válidas de IA generativa neste contexto. Sua inclusão permite uma compreensão histórica e abrangente das aplicações de IA na mediação de taxonomias.

A partir da análise sistemática observou-se que os artigos se distribuem em quatro grandes enfoques metodológicos principais, que definem o papel da IA generativa no ciclo de vida das taxonomias:

1. *Construção ou indução de taxonomias a partir de dados textuais*

Esse grupo reúne estudos, como os de Chen e Varró (2023) e Moraes *et al*. (2024), que utilizam modelos de linguagem para a geração automática de hierarquias taxonômicas, geralmente a partir de coleções textuais não estruturadas. A IA generativa é empregada para prever relações hiperônimas com base em prompts semânticos, estruturar árvores hierárquicas e classificar pares de conceitos quanto à existência de vínculos taxonômicos. Destacam-se, nessas abordagens, o uso de LLMs em cenários zero-shot, a exploração de representações vetoriais não supervisionadas e a combinação entre geração e ranqueamento por perplexidade.

2. *Expansão e enriquecimento de taxonomias existentes*

Nos estudos dessa categoria, a IA generativa é utilizada para propor novos termos ou sugerir inserções hierárquicas em taxonomias preexistentes. Os métodos adotados variam entre o uso de *embeddings* semânticos para aproximação vetorial entre termos (sem geração textual explícita) e abordagens baseadas em modelos seq2seq, voltadas à geração autônoma de nomes de conceitos. Alguns estudos combinam automação e validação humana, especialmente quando aplicados a taxonomias de alta relevância institucional. Um exemplo é o estudo de D'Amico *et al.* (2024), que propõe o enriquecimento automático da taxonomia de competências digitais da ESCO, incorporando novos termos extraídos da web.

3. *Avaliação e refinamento automatizado de taxonomias*

Diversos estudos investigam o uso de modelos generativos como avaliadores da coerência semântica e estrutural de taxonomias, substituindo ou complementando métodos estatísticos tradicionais. Modelos como GPT-4, Claude-3 e Qwen-14B são empregados para estimar a adequação de tópicos, identificar duplicatas ou outliers e sugerir refinamentos estruturais, como a eliminação de ciclos, a unificação de sinônimos e a reestruturação de vínculos taxonômicos. Essa linha de pesquisa destaca-se pelo uso de *prompting* com raciocínio explícito, inspirado na técnica *Chain-of-Thought*, além do desenvolvimento de pipelines totalmente automatizados, como no caso do projeto WiKC, aplicado à Wikidata.

4. *Geração de conceitos e rótulos taxonômicos*

Embora menos frequente na literatura revisada, essa abordagem é também uma das mais inovadoras, pois foca na geração autônoma de conceitos linguísticos a partir de contextos hierárquicos. Utilizando modelos como T5, GRU e seq2seq baseados em grafos, os sistemas são capazes de nomear nós intermediários ausentes na estrutura taxonômica com base em informações distribuídas entre conceitos vizinhos. Os estudos analisados indicam que a IA generativa pode produzir termos compostos coerentes, inclusive em domínios técnicos e especializados. No entanto, a validação humana ainda é essencial para garantir a precisão e a adequação dos resultados.

A partir da análise dos enfoques metodológicos, algumas tendências recorrentes na literatura podem ser identificadas. Destaca-se a predominância do uso de modelos de linguagem de larga escala (LLMs), tanto comerciais (GPT-4, Gemini, Copilot e outros) quanto *open source* (LLaMA, Mistral). Observa-se também uma variação entre o uso direto (geração textual) e indireto (*embeddings*), bem como a ampla adoção de *prompting* estruturado e instruções *zero-shot*.

Por fim, nota-se que nenhum dos estudos analisados aborda explicitamente o papel de bibliotecários, taxonomistas ou demais profissionais da organização da informação na curadoria dos resultados gerados por sistemas de IA. Essa lacuna representa uma oportunidade relevante de investigação, especialmente no que diz respeito à colaboração entre inteligência artificial e especialistas humanos para a validação, refinamento e aplicação de taxonomias em contextos informacionais.

Fase 3 Análise dos Métodos e Técnicas de IA Generativa Aplicadas as Taxonomias

Na última fase da pesquisa, os documentos recuperados e selecionados foram analisados com o objetivo de compreender de que maneira as abordagens de IA são aplicadas às taxonomias destacando potencialidades, desafios e oportunidades.

A aplicação de IA generativa no desenvolvimento e manutenção de taxonomias representa um avanço significativo, permitindo automação e escalabilidade. No entanto, a incorporação dessas tecnologias traz desafios metodológicos e conceituais, conforme quadro 1, que precisam ser analisados para garantir que os modelos não comprometam a coerência e a qualidade das estruturas taxonômicas.

Quadro 1. *Aplicação de IA generativa em taxonomias*

Uso de IA generativa aplicada à taxonomia	*Potencialidades*	*Desafios*	*Exemplos de artigos analisados*
Geração automática de conceitos e relações hierárquicas	A IA gera novos nós, nomes e relações parônimas com fluidez semântica e linguística.	Risco de incoerência semântica, tal como conceituações ambíguas, termos genéricos ou relações inadequadas podem comprometer a integridade da taxonomia.	Artigo 1: Enhancing taxonomy completion with conception generration via fusing relational representations (Zeng *et al*., 2021). Artigo 2: Taxonomy completion via implicit concept insertion (Shi *et al*., 2024).
Enriquecimento com linguagem natural contextualizada	Inclusão de termos emergentes a partir de descrições coletadas da web e de usuários.	Introdução de ruído semântico, a informalidade, ambiguidade ou superficialidade dos dados textuais podem degradar a qualidade conceitual da estrutura.	Artigo: Taxonomy enrichment: a framework for automatic updates and labor market analysis (Santo, 2024).
Reformulação estrutural automatizada	Construção de sistemas que refinam árvores complexas, removem redundâncias e reorganizam hierarquias.	Ausência de justificativa das mudanças, impactam na rastreabilidade e visibilidade de informações claras sobre as transformações aplicadas. Acarretando a dificuldade de corrigir e auditar as taxonomias.	Artigo: Refining Wikidata taxonomy using large language models (Peng *et al*., 2024)
Inserção de conceitos implícitos e intermediários	Preenchimento de lacunas não explícitas, fortalecendo a coesão textual.	Modelos podem inserir conceitos não validados ou irrelevantes, criando ruído ou duplicidade na taxonomia.	Artigo: Taxonomy completion via implicit concept insertion (Shi *et al*., 2024).
Adaptação dinâmica ao contexto	Taxonomias são moldadas conforme domínio, tempo, local ou perfil de uso.	Múltiplas versões simultâneas podem afetar a interoperabilidade e a governança da taxonomia.	Artigo: Using Zero-shot prompting in the automatic creation and expansión of topic taxonomies for taggin retail banking transactions (Moraes *et al*., 2024).
Generalização para diferentes domínios e atividades	Modelos atuam em diversos contextos sem reconfiguração, inclusive em zero-shot.	A versatilidade empregada pelos modelos pode comprometer a profundidade semântica e a precisão em domínios especializados.	Artigo: TaxoLLaMA: WordNet-based model for solving multiple lexical semantic tasks (Moskvoretskii *et al*., 2024).
Escalabilidade e atualização contínua	Capacidade de operar em fluxos contínuos de dados e atualizar estruturas taxonômicas.	Atualizações automatizadas podem sobrepor decisões humanas bem fundamentadas, gerando instabilidade.	Artigo: Refining Wikidata taxonomy using large language models (Peng *et al*., 2024).
Integração com sistemas de recomendação e recuperação da informação	Taxonomias geradas por IA otimizam buscas e categorização.	Estruturas adaptadas por IA podem reforçar vieses e comprometer a neutralidade da organização do conhecimento.	Artigo: TacoPrompt: a collaborative multi-task prompt learning method for self-supervised taxonomy completion (Xu *et al*., 2023).

A análise evidencia que, embora a IA generativa traga ganhos em automação e escalabilidade, sua implementação em taxonomias requer estratégias de validação robustas. O equilíbrio entre automação e controle humano surge como um fator crítico para garantir qualidade, coerência e aplicabilidade das taxonomias geradas. Pesquisas futuras podem explorar mecanismos híbridos, combinando IA generativa com curadoria humana para aprimorar a confiabilidade dos sistemas taxonômicos.

Por último, no que tange aos caminhos para inovação, os estudos analisados apontam oportunidades relevantes para instituições e pesquisadores. Destaca-se o uso da IA no suporte à curadoria de repositórios colaborativos, como demonstrado pelas soluções TaxoLLaMA e WiKC, que auxiliam na verificação, padronização e enriquecimento contínuo de plataformas como a Wikidata. Também se evidenciam aplicações integradas a sistemas de busca e recomendação, em que taxonomias enriquecidas por IA ampliam a precisão e a relevância dos resultados oferecidos aos usuários.

5 Considerações finais

Este estudo investigou o uso da Inteligência Artificial (IA) generativa na otimização da construção e reformulação de taxonomias, destacando suas potencialidades e desafios. A pesquisa evidenciou que a IA generativa pode automatizar a construção de taxonomias, oferecendo técnicas e métodos eficientes para a organização e classificação de grandes volumes de dados. Além disso, a IA demonstrou capacidade para identificar padrões semânticos, sugerir novos termos e relações hierárquicas, e enriquecer taxonomias existentes com termos emergentes extraídos de fontes dinâmicas, como a web.

Entre as potencialidades identificadas, destaca-se a geração automática de conceitos e relações hierárquicas, o enriquecimento com linguagem natural contextualizada e a reformulação estrutural automatizada das taxonomias. Esses avanços permitem maior flexibilidade e escalabilidade na gestão de grandes volumes de informação, além de otimizar a adaptação das taxonomias às mudanças dinâmicas no ambiente organizacional.

No entanto, a pesquisa também revelou desafios importantes na aplicação da IA generativa para a reformulação de taxonomias. A confiabilidade das classificações automáticas, a qualidade dos dados, a presença de vieses nos algoritmos, e as questões de ambiguidade semântica e falta de contexto são obstáculos que podem comprometer a precisão e a integridade das taxonomias geradas. Além disso, a interação humana continua sendo crucial, principalmente na validação das modificações e na supervisão da curadoria dos resultados, especialmente em domínios especializados.

A pesquisa também apontou para uma lacuna importante na literatura sobre o papel dos profissionais da Ciência da Informação, como bibliotecários e taxonomistas, na curadoria dos resultados gerados por sistemas de IA. Essa lacuna representa uma oportunidade significativa para futuras investigações, que podem explorar formas de colaboração entre inteligência artificial e especialistas humanos, visando garantir a qualidade, a confiabilidade e a relevância das taxonomias aplicadas.

Em suma, embora a IA generativa apresenta um grande potencial para otimizar a construção e reformulação de taxonomias, sua implementação bem-sucedida exige um equilíbrio entre a tecnologia e a expertise humana. A colaboração entre as capacidades computacionais da IA e o conhecimento especializado dos profissionais da informação é essencial para garantir que as taxonomias geradas sejam precisas, relevantes e adaptáveis às necessidades específicas das organizações. Este estudo, portanto, contribui para a compreensão do papel da IA generativa na Ciência da Informação e abre novas possibilidades para a pesquisa e inovação nesse campo.

Referências

ADRIANI, M. L. *Um método para a construção de taxonomias utilizando a DBPEDIA*. 149 fls. 2017. Dissertação (Mestrado em Engenharia do Conhecimento) – Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2017.

AGANETTE, E. C.; MACULAN, B. C. M. S. *The complexity of the classification*: represatantion in information Science education. [*S.l.*]: Seven Editora, 2023. Disponível em: https://sevenpublicacoes.com.br/editora/article/view/2346. Acesso em: 24 fev. 2025.

ANDERSON, L. W.; KRATHWOHL, D. R. *A taxonomy for learning, teaching, and assessing*: A revision of Bloom's taxonomy of educational objectives. [*S.l.*]: Longman, 2001. Disponível em: https://www.worldcat.org/title/taxonomy-for-learning-teaching-and-assessing. Acesso em: 24 fev. 2025.

BAILEY, K. D. *Typologies and Taxonomies*: An Introduction to Classification Techniques. [*S.l.*]: SAGE Publications, 2012. Disponível em: https://us.sagepub.com/en-us/nam/typologies-and-taxonomies/book52363. Acesso em: 24 fev. 2025.

CHEN, B; FANDI; VARRÓ, D. Prompting or Fine-tuning? A Comparative Study of Large Language Models for Taxonomy Construction. 2023. *In:* ACM/IEEE International Conference on Model Driven Engineering Languages and Systems Companion (MODELS-C), Västerås, Sweden, 2023. *Proceedings [...]*. Västerås: IEEE, 2023. pp. 588-596.

BLOOM, B. S. *Taxonomy of Educational Objectives*: The Classification of Educational Goals. [*S.l.*]: Longman, 1956. Disponível em: https://eric.ed.gov/?id=ED056243. Acesso em: 24 fev. 2025.

CASAR CORREDERA, J. R. Inteligencia artificial generativa. *Anales de la Real Academia de Doctores de España*, Madrid, v. 8, n. 3, p. 475-489, 2023. Disponível em: https://www.rade.es/imageslib/PUBLICACIONES/ARTICULOS/V8N3%20-%2001%20-%20ED%20-%20CASAR.pdf. Acesso em: 8 fev. 2025.

COZMAN, F. G.; NERI, H. O que, afinal, é inteligência artificial? *In:* COZMAN, F. G.; PLONSKI, G. A.; NERI, H. (org.) *Inteligência artificial*: avanços e tendências. São Paulo: Instituto dos estudos avançados, 2021. Disponível em: https://www.mpsp.mp.br/portal/page/portal/documentacao_e_divulgacao/doc_biblioteca/bibli_servicos_produtos/BibliotecaDigital/BibDigitalLivros/TodosOsLivros/Inteligencia-artificial=avancos-e-tendencias.pdf. Acesso em: 12 fev. 2025.

D'AMICO, S. *et al*. Enriching Skill Taxonomies through Vector Space Models. 2024. *In*: IEEE International Conference on Big Data (BigData), Washington, 2024. *Proceedings [...]*. Washington: IEEE, 2024. pp. 2297-2302. DOI: 10.1109/BigData62323.2024.10825415. Acesso em: 03 mar. 2025.

GONZÁLEZ, R.; PEREIRA, J. Machine Learning Applications in Taxonomy Development. *Information Systems Review*, [*S.l.*], 2022. Disponível em: https://www.infosysreview.com/articles/machine-learning-taxonomy. Acesso em: 14 mar. 2025.

GRUBER, T. R. A translation approach to portable ontology specifications. *Knowledge Acquisition*, [*S.l.*], v. 5, n. 2, p. 199-220, 1993. Disponível em: https://tomgruber.org/writing/ontology-approach.pdf. Acesso em: 24 fev. 2025.

MAYR, E. *Principles of Systematic Zoology*. New York: McGraw-Hill, 1969. Disponível em: https://archive.org/details/principlesofsystematiczoology. Acesso em: 24 fev. 2025.

MCCARTHY, J. *Formalizing Common Sense: Papers by John McCarthy*. Norwood: Ablex Publishing, 1995. Disponível em: https://www-formal.stanford.edu/jmc/common-sense.html. Acesso em: 24 fev. 2025.

MORAES, D. *et al.* Using Zero-shot Prompting in the Automatic Creation and Expansion of Topic Taxonomies for Tagging Retail Banking Transactions. *arXiv*, [*S.l.*], v. 1, jan. 2024. Disponível em: https://arxiv.org/abs/2401.06790. Acesso em: 20 mar. 2025.

MOSKVORETSKII, V. *et al.* TaxoLLaMa: WordNet-based model for solving multiple lexical semantic tasks. *In*: ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 62., 2024, Bangkok. *Proceeding […]*. New York: Arxiv, 2024. Disponível em: https://arxiv.org/pdf/2403.09207. Acesso em: 29 abr. 2025.

NICKERSON, R. C.; MUNTERMANN, J.; VARSHNEY, U. Taxonomy Development in Information Systems: Developing a Taxonomy of Mobile Applications. *In:* European Conference on Information Systems (ECIS), 2009. *Proceedings [...]*. [*S.l.*]: ECIS, 2009. Disponível em: https://aisel.aisnet.org/ecis2009/158/. Acesso em: 25 mar. 2025.

PENG, Y. *et al*. Refining Wikidata taxonomy using large language models. *In:* ACM INTERNATIONAL CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT, 33, 2024, Boise. *Proceedings [...]*. New York: ACM, 2024. Disponível em: https://dl.acm.org/doi/10.1145/3627673.3679156. Acesso em: 29 abr. 2025.

PRODANOV, C. C.; FREITAS, E. C. *Metodologia do trabalho científico*. 2. ed. Novo Hamburgo: Feevale, 2013.

RANGANATHAN, S. R. *Colon Classification*. [*S.l.*]: Madras Library Association, 1933. Disponível em: https://www.isko.org/cyclo/colon-classification. Acesso em: 26 fev. 2025.

ROWLEY, J. *Organizing Knowledge*: An Introduction to Information Retrieval. [*S.l.*]: Gower Publishing, 1992. Disponível em: https://www.worldcat.org/title/organizing-knowledge-an-introduction-to-information-retrieval. Acesso em: 26 fev. 2025.

RUSSELL, S; NORVIG, P. *Artificial inteligence*: a modern approach. 4. ed. Hoboken: Pearson, 2021.

SANTO, A. de. Taxonomy enrichment: a framewok for automatic updates and labor market analysis. *In:* INTERNATIONAL CCONFERENCE OF THE ITALIAN ASSOCIATION OR ARTIFICIAL INTELLIGENCE BOLZANO, 23, 2024, Italy. *Proceedings [...]*. Italy: [*S. n.;*], 2024. Disponível em: https://ceur-ws.org/Vol-3914/short71.pdf. Acesso em: 29 abr. 2025.

SHI, J. *et al.* Taxonomy completion via implicit concept insertion. *In:* ACM WEB CONFERENCE, 2024, Singapore. *Proceeding […]*. New York: ACM, 2024. Disponível em: https://dl.acm.org/doi/10.1145/3589334.3645584. Acesso em: 29 abr. 2025.

SHIRKY, C. *Ontology is Overrated*: Categories, Links, and Tags. [*S.l.; S. n.;*], 2005. Disponível em: http://www.shirky.com/writings/ontology_overrated.html. Acesso em: 24 fev. 2025.

SMITH, J.; JOHNSON, L. Dynamic Taxonomies in the Age of Big Data. *Data Science Review*, [*S.l.*], 2023. Disponível em: https://www.datasciencereview.com/articles/dynamic-taxonomies-big-data. Acesso em: 18 mar. 2025.

TREVISAN, L. C; DAMIAN, I. P. M. Gestão do conhecimento: diretrizes e práticas recomendadas às organizações. *Ciência da Informação*, Brasília, v. 47, n. 2, p.21-34, maio/ago. 2018. Disponível em: https://www.brapci.inf.br/v/99232. Acesso em: 17 fev. 2025.

XU, H. *et al.* TacoPrompt: a collaborative multi-task prompt learning method for self-fupervised taxonomy completion. *In:* CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING, 2023, Singapore. *Proceedings [...]*. Singapore: ACL, 2023. Disponível em: https://aclanthology.org/2023.emnlp-main.979.pdf. Acesso em: 29 abr. 2025.

ZENG, Q. *et al.* Enhancing taxonomy completion with concept generation via fusing relational representations. *In*: KDD, 2021, Singapore. *Proceedings [...]*. Sinagore: KDD, 2021. Disponível em: https://arxiv.org/abs/2106.02974. Acesso em: 29 abr. 2025.

WOODS, E. The corporate taxonomy: creating a new order. *KMWorld*, Camden, v.13, n. 7, July/Aug. 2004.